



NeuralStroke

NEURALSTROKE

NTI Final Project



OUR TEAM

- Mohammed Montaser
- Ahmed Dahy
- Merna Medhat
- Karen Medhat

UNDER THE SUPERVISION OF

Eng. Aya Saleh



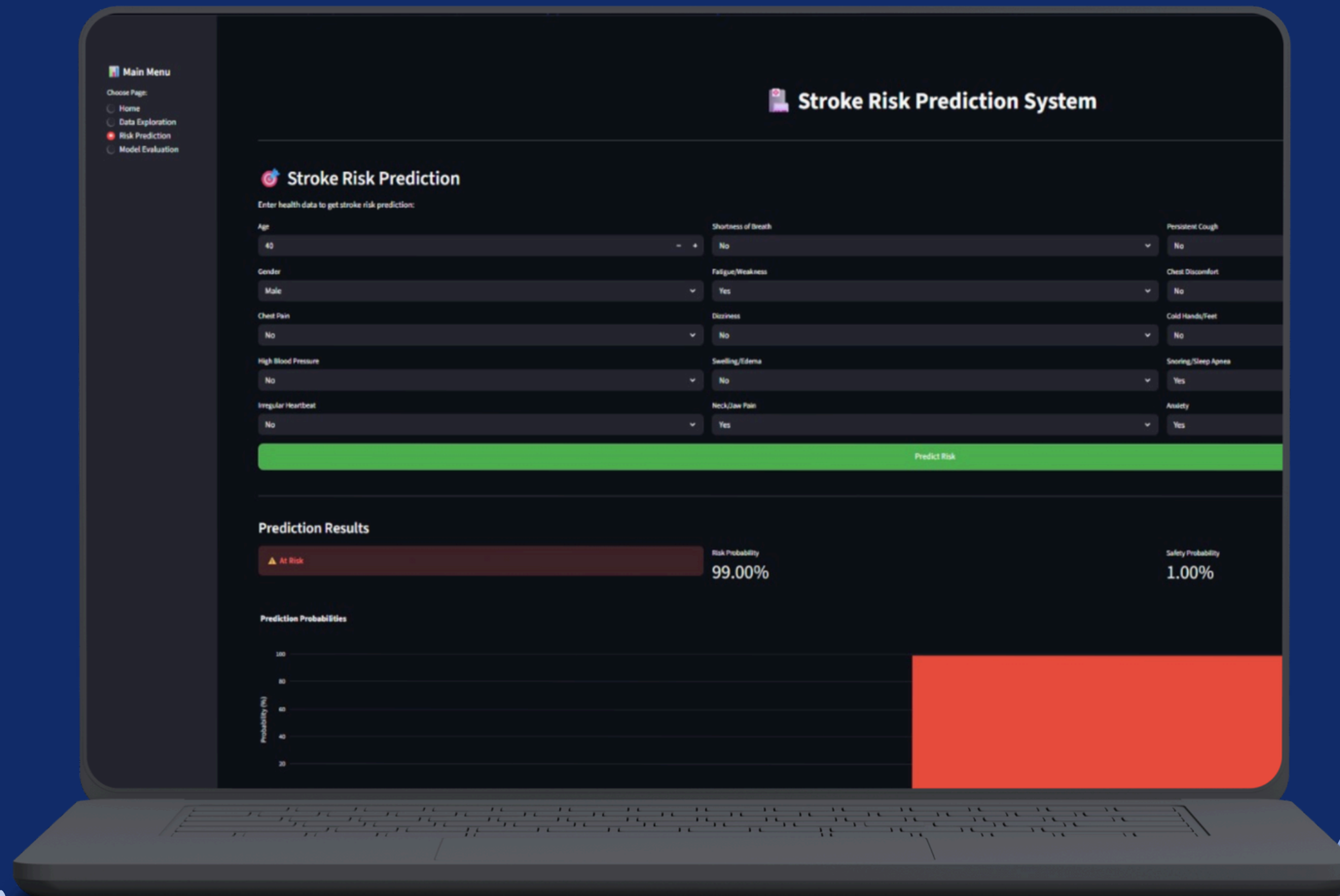
LIST OF CONTENT

- Problem Statement
 - Our Solution
 - The Dataset
 - Preprocessing Steps
 - Feature Engineering
 - Our Models & Accuracy
- 

PROBLEM STATEMENT

Cardiovascular diseases, specifically strokes, remain a leading cause of global mortality and long-term disability. Despite the presence of identifiable warning signs, early detection is often hindered by the complex, non-linear relationship between age, gender-specific risk factors, and symptomatic precursors.

OUR SOLUTION



OUR SOLUTION

- Developed automated risk screening system.
- Leverages advanced predictive analytics.
- Models complex biological risk patterns.
- Real-time classification for at-risk patients.
- Continuous learning from clinical updates.
- Bridging data science and medicine.






OUR SOLUTION

- Integrated 35,000 medically validated samples.
- Applied sigmoidal age-risk mathematical functions.
- Factored gender-specific metabolic risk multipliers.
- Removed irrelevant features via RFE.
- Validated models against medical literature.
- Provided explainable results for doctors.





THE DATASET

- 35,000 high-fidelity medical samples.
 - Balanced: 61.6% at-risk, 38.4% healthy.
 - 16 features: symptoms and demographics.
 - Binary and regression targets included.
 - Large sample size ensures generalizability.
- 



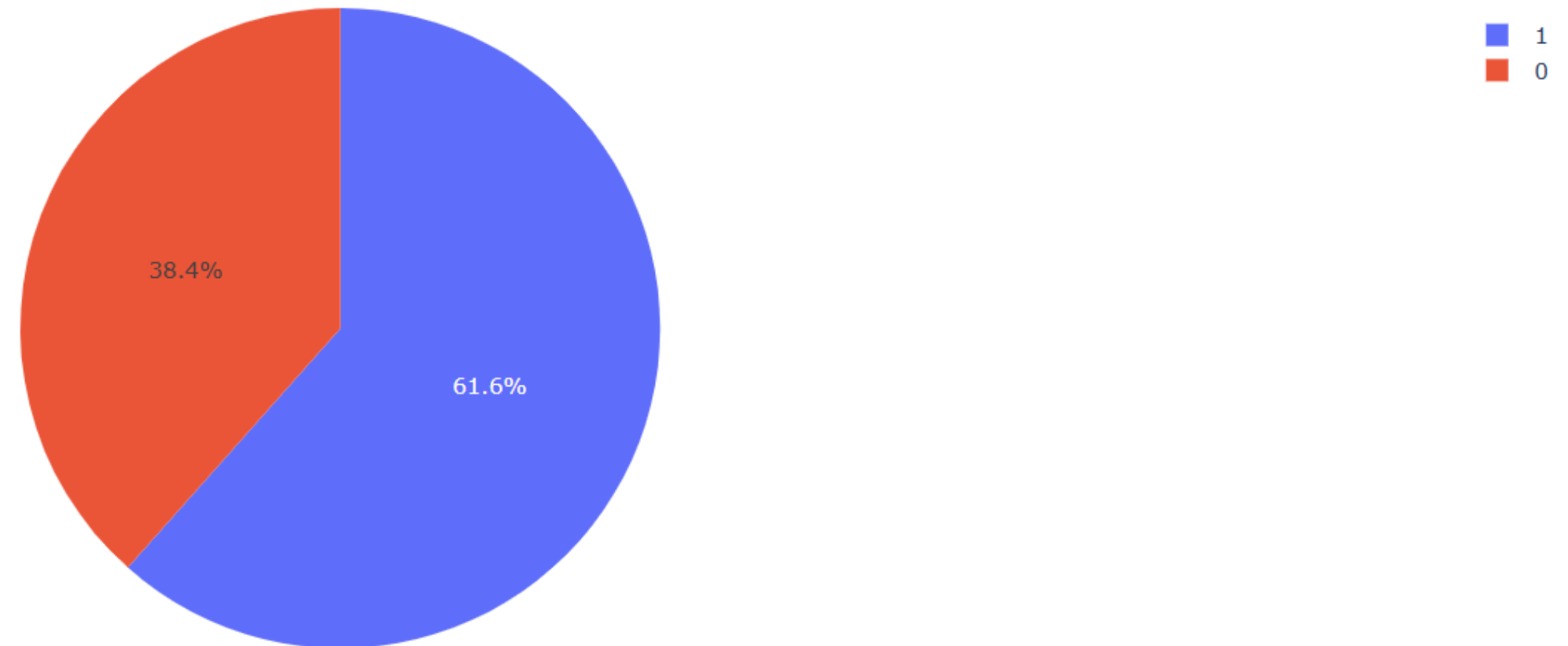
PREPROCESSING STEPS

- Conducted initial raw data cleaning.
- Removed duplicated patient records.
- Encoded gender and categorical variables.
- Verified zero null value integrity.
- Split data into 70/30 sets.



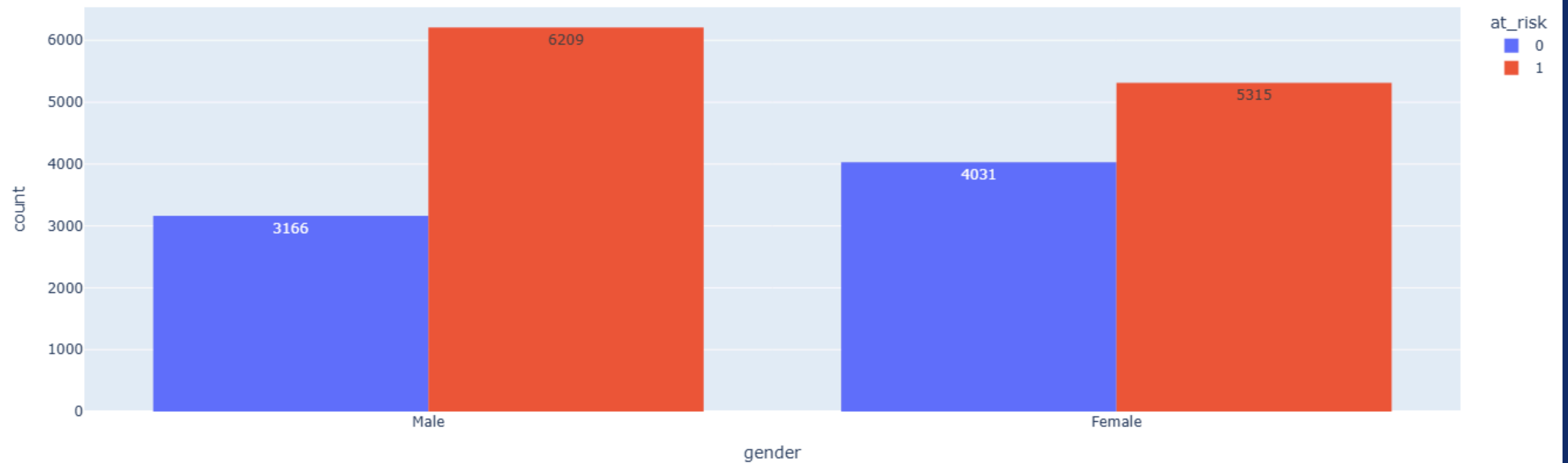
EDA RISK

at_risk Distribution

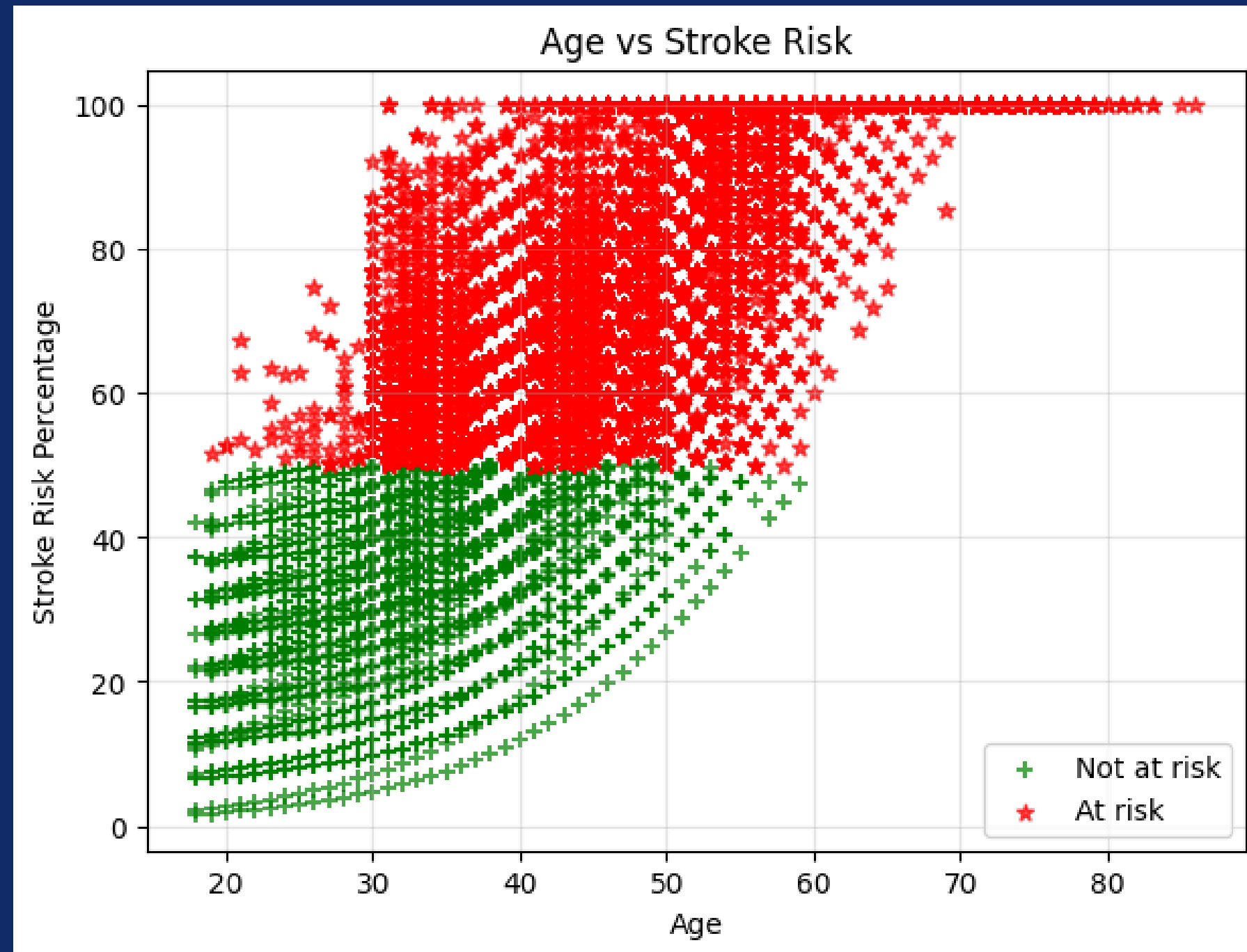


EDA GENDER RISK

Distribution of Gender by Risk Status

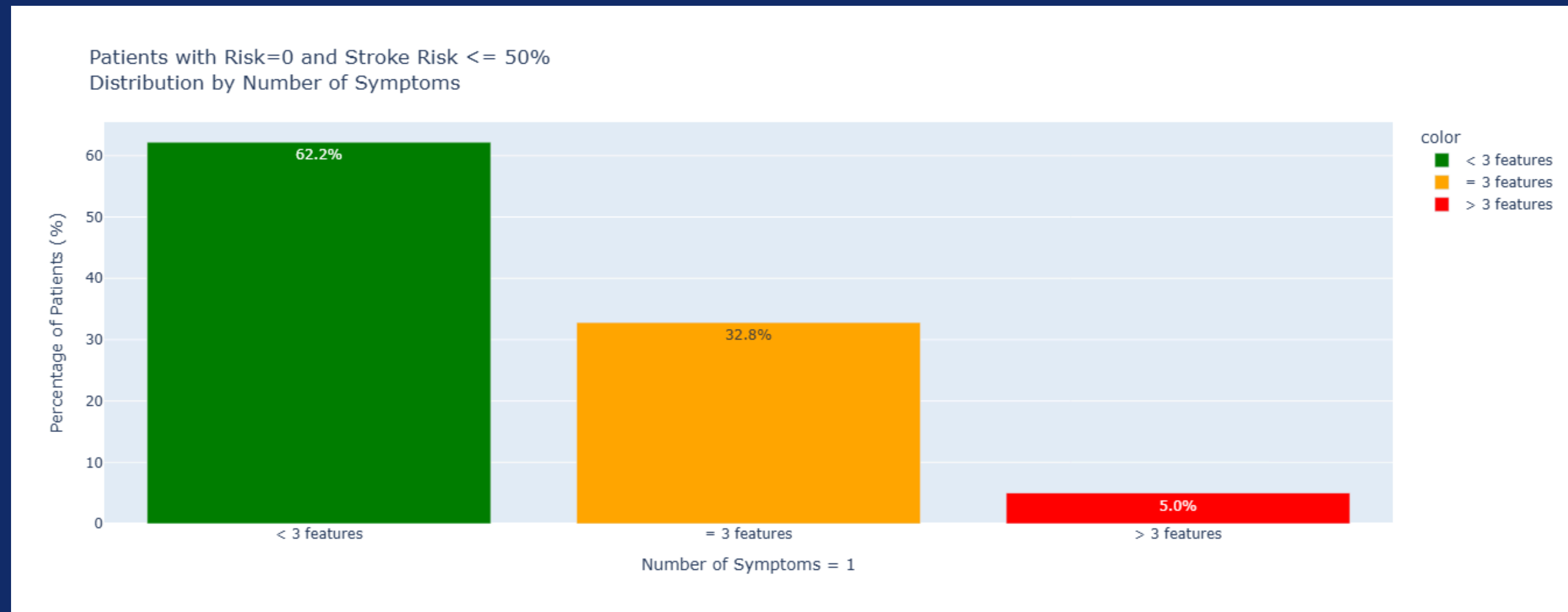


EDA AGE STROKE RISK PERCENTAGE



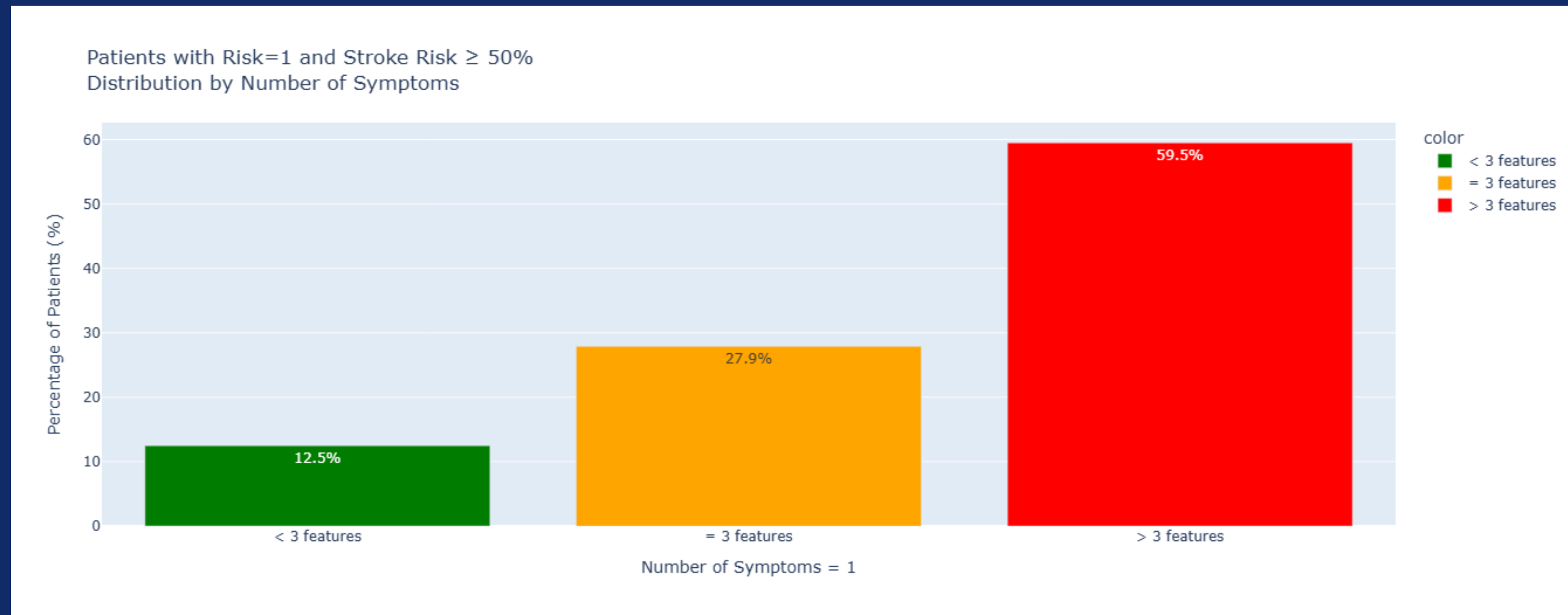
ANALYSIS OF SYMPTOM COUNT DISTRIBUTION

PATIENTS WITH RISK=0 AND STROKE RISK \leq 50

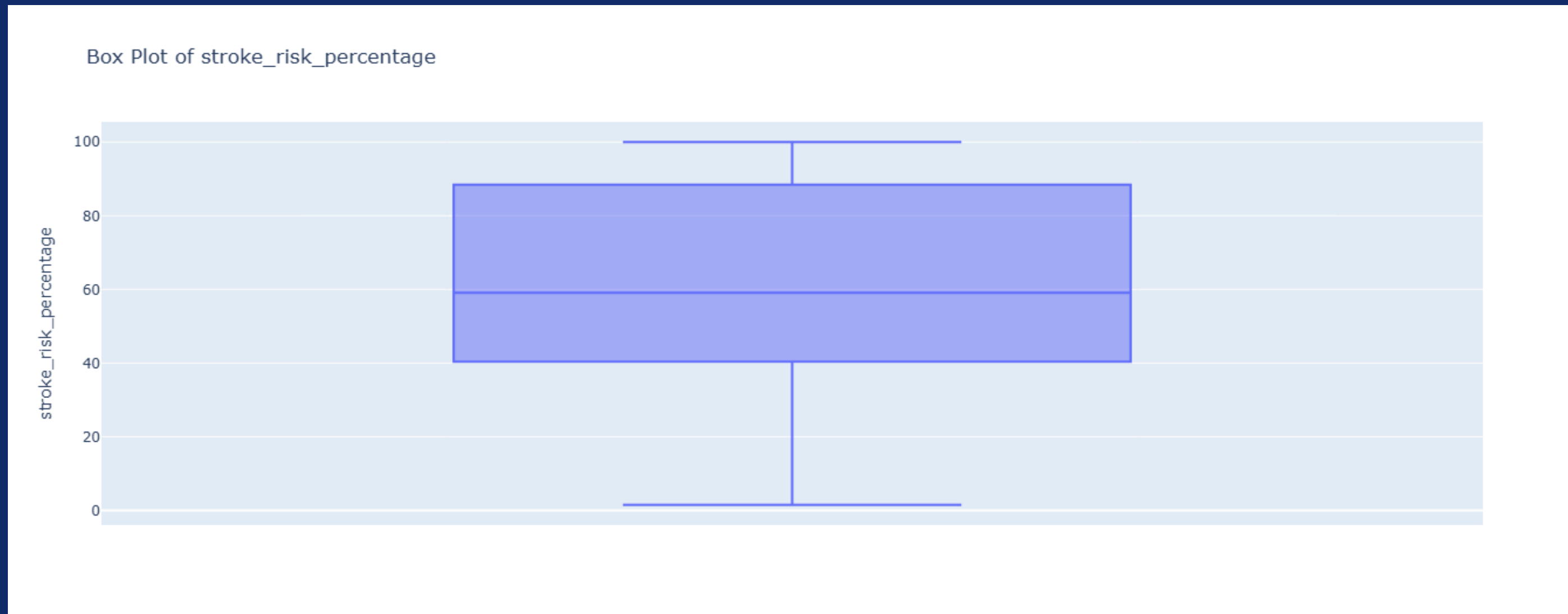


ANALYSIS OF SYMPTOM COUNT DISTRIBUTION

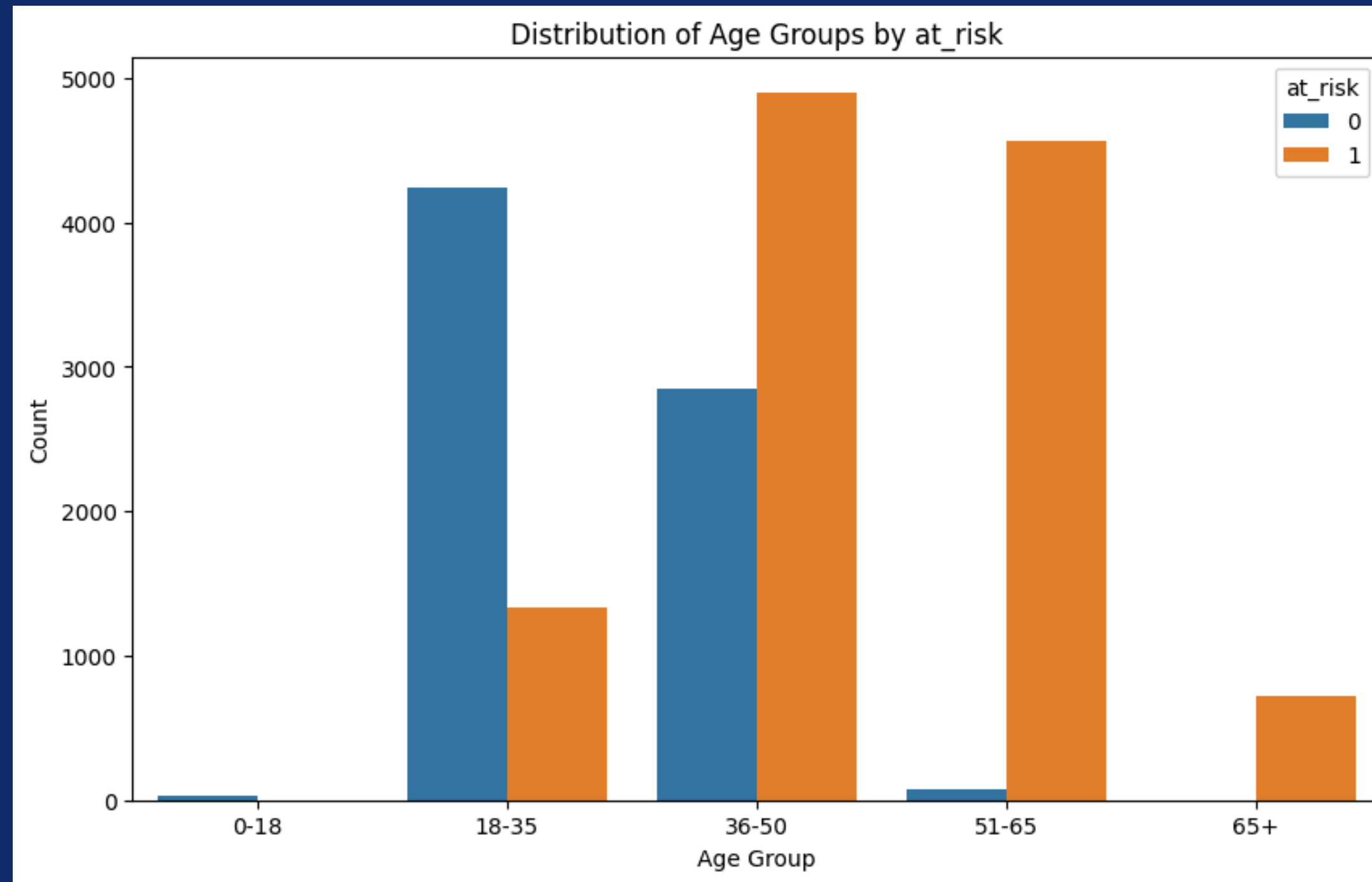
PATIENTS WITH RISK=1 AND STROKE RISK ≥ 50



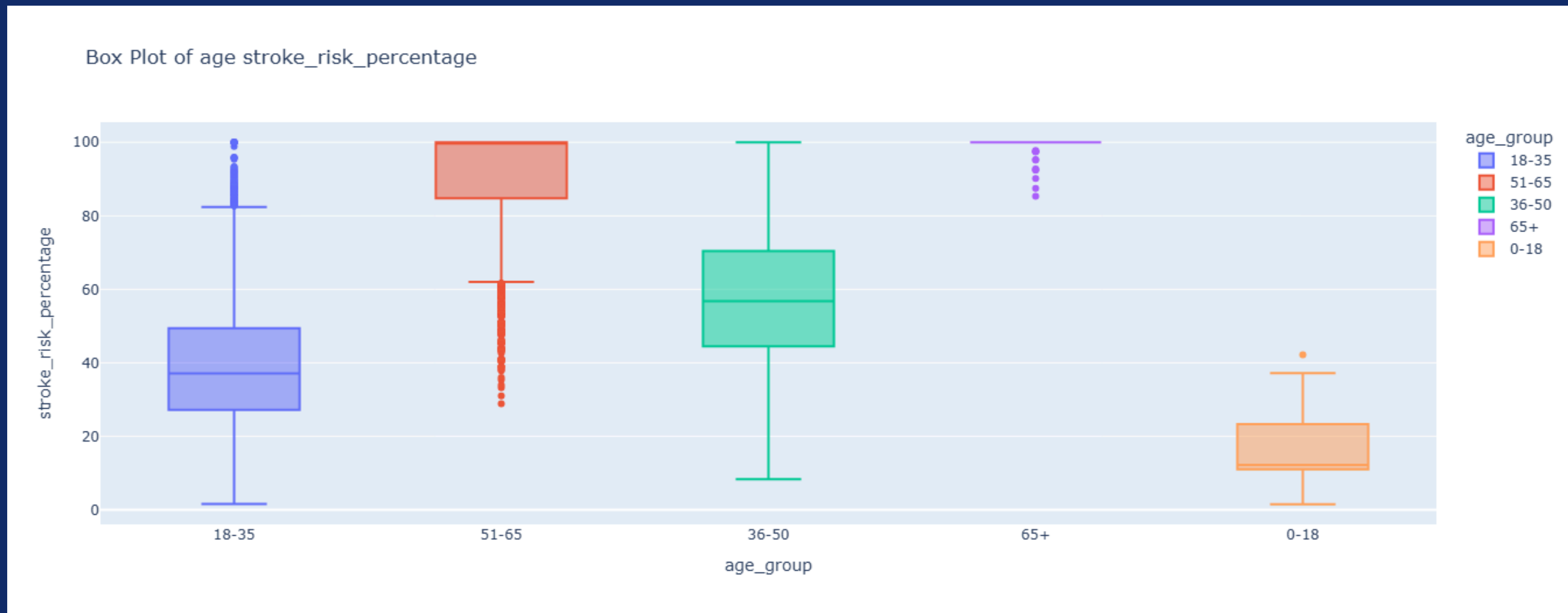
BOXPLOT STROKE RISK PERCENTAGE



EDA AGE GROUP RISK



BOXPLOT AGE GROUP STROCK PERCENTAGE





FEATURE ENGINEERING

- Performed deep Correlation Matrix analysis.
- Identified multicollinear features.
- Used Recursive Feature Elimination (RFE).
- Dropped weak predictors like nausea.





OUR MODELS & ACCURACY

- **Logistic Regression:** established statistical baseline.
- **SVC:** optimized high-dimensional decision boundaries.
- **Random Forest:** ensemble of trees.
- **AdaBoost:** sequential boosting for precision.
- **Gradient Boosting:** our highest-performing algorithm.
- **Deep Learning:** multi-layer Neural Network.



OUR MODELS & ACCURACY

- **Gradient Boosting:** 94.7% peak accuracy.
- **Random Forest:** 93.9% peak accuracy.
- **SVM:** 96.8% very high precision.
- **Logistic Regression:** 96.78% ROC Accuracy.
- **Neural Network:** strong convergence, %96.10 Accuracy.
- Validated using ROC-AUC and Confusion.




FUTURE WORKS

- Integrate real-time biometric vital signs.
 - Add dynamic heart rate variability.
 - Track longitudinal symptom progression data.
 - Expand features to include comorbidities.
 - Validate across diverse global demographics.
 - Incorporate patient lifestyle and activity.
- 



FUTURE WORKS

- Develop mobile risk screening application.
 - Create interactive doctor-patient XAI dashboards.
 - Benchmark medical data generative models.
 - Secure cloud-based medical API integration.
 - Implement system in hospital triage.
 - Conduct large-scale clinical validation trials.
- 



THANK YOU